



به نام هستی بخش  
"و سلام بر مهدی که انتظارش را نه فقط دل عاشق،  
که ترنم هر باران بهاری و هر روزنه‌ی امید میکشد..."  
**تمرین 9 (فصل 14، 15)**  
درس پایگاه داده‌ها، بهار ۱۴۰۳

1) به سوالات زیر پاسخ دهید: (20 نمره)

الف) روش‌های Sort-Merge Join و Hash Join را با یکدیگر مقایسه کرده و توضیح دهید هر کدام در چه شرایطی بهتر عمل می‌کنند.

ب) چهار نوع استفاده از شاخص در پرس‌وجوهای تک رابطه‌ای را نام برده و هر کدام را توضیح دهید.

ج) به چه دلایلی ممکن است برای خواندن داده‌ها به سراغ روش full table scan برویم؟

2) فرضیات زیر را در رابطه با یک پایگاه داده در نظر بگیرید و هزینه پرس‌وجوهای داده شده را به لحاظ عملیات I/O روی دیسک بدست آورید. (30 نمره)

- یک رابطه  $R(A, B, C)$  با 100,000 تاپل وجود دارد.
- هر تاپل رابطه  $R$  به اندازه 100 بایت است.
- یک رابطه دیگر  $S(A, D)$  با 50,000 تاپل وجود دارد.
- هر تاپل رابطه  $S$  به اندازه 80 بایت است.
- پایگاه داده از اندازه صفحه 4 کیلوبایت (4096 بایت) استفاده می‌کند.
- اندازه صفحه برای هر دو رابطه 4 کیلوبایت است.
- 20 صفحه بافر برای پرس‌وجوها موجود است.

الف) انجام یک اسکن کامل جدول روی رابطه  $R$

ب) انجام Sort-Merge Join بین رابطه  $R$  و رابطه  $S$  بر روی صفت  $A$

ج) انجام Hash Join بین رابطه  $R$  و رابطه  $S$  بر روی صفت  $A$

(3) روابط زیر را در نظر بگیرید. (25 نمره)

Customer (ssn, name, occupation, age); 24,000 tuples, 300 tuples/page

Rent (ssn, gmodel, duration); 180,000 tuples, 400 tuples/page

GPU (model, size, memtype, cores); 12,000 tuples, 300 tuples/page

- صفت occupation در رابطه Customer ، میتواند 20 مقدار مختلف داشته باشد. (با توزیع uniform )
- صفت memtype در رابطه GPU، میتواند 10 مقدار مختلف داشته باشد. (با توزیع uniform )
- طول همه مشخصه ها باهم برابر است

- 10 صفحه بافر در اختیار داریم
- یک شاخص درختی unclustered با ارتفاع 3 روی Customer.ssn داریم

هزینه پرس و جو زیر را با استفاده از plan روبرو تخمین بزنید  $((G \bowtie (SMJ)R) \bowtie (INLJ)C)$

```
SELECT R.gmodel, avg(R.duration)
```

```
FROM Customer C, GPUs G, Rent R
```

```
WHERE C.ssn = R.ssn AND R.gmodel = G.model AND C.occupation = 'Dev' AND G.memtype =  
'GDDR6X' GROUP BY R.gmodel
```

انتظار می رود selection ها و projection ها را به شکلی در نظر بگیرید که کمترین هزینه را داشته باشند. Plan پیشنهادی خود را رسم کنید.

(4) روابط زیر را در نظر بگیرید: (۲۵ نمره)

Item(Name,Category)

Store(Name, City, StreerAddr)

Transaction(ItemName, StoreName, Date)

با استفاده از اطلاعاتی که در ادامه آمده است، روش بهینه برای پرس و جوی زیر را به همراه هزینه آن بدست آورید. فرض کنید 50 صفحه بافر در اختیار داریم.

Item: 50,000 tuples, 10 tuples/page

Index: Unclustered hash on Name

Store: 1,000 tuples, 5 tuples/page; 100 cities

Index1: Unclustered hash index on Name

Index2: Clustered B+ tree (Alt.2) on City

Transaction: 500,000 tuples, 25 tuples/page; 10 items bought per store per day. The relation stores transactions committed over a 50 day period

Index: clustered B+ tree (Alt.2) on the pair of attributes StoreName, Date

```
SELECT I.Category, S.City  
FROM Item I, Store S, Transaction T  
WHERE I.Name = T.ItemName AND T.StoreName = S.Name  
AND T.Date = '2004-12-15' AND S.City = 'NYC'
```

سر بلند و پیروز باشید.